



August 14, 2000

DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-05r1

MEMORANDUM FOR Charisse Jones
Leader, A.C.E. Computer Processing Team

From: Danny R. Childers *DR*
Leader, A.C.E. Design Team

Subject: Specifications for Computer Preprocessing for Missing Data Input
Files for the Accuracy and Coverage Evaluation (A.C.E.)
Revision 1

Prepared by: Michael Beaghen

This memorandum specifies computer coding of the A.C.E. person data from the person match and review coding system (PerMaRCS) after the A.C.E. matching is complete, but preliminary to the creation of the missing data P-sample and E-sample input files. In particular, the variables TESPERS and HICODE are recoded, and three derived variables, WHHNNGP, WPFIN and STRCDE are created. The variable TESPERS indicates whether a person is a targeted extended search (TES) person. While a TES person flag is maintained in the PerMaRCS, this flag is not set for census people in relist and list/enumerate clusters and requires setting. Furthermore, we want to change people flagged in PerMaRCS as TES people who are duplicates or did not have sufficient information for matching to non-TES person status. The second variable to be recoded is the HICODE. Census people in housing units that did not have a TES field followup done and are in clusters where TES matching occurs will be recoded from 'GE' to 'GU' to facilitate a later imputation for them. The recoding of HICODE should occur after the recoding of TESPERS.

The first derived variable, WHHNNGP, is created to indicate whether the household is coded as a whole household group quarters/other residence or if the household is a whole household geocoding error. The second derived variable, WPFIN, is created to indicate whether the household is coded as a whole household match, non-match, or partial household match. Lastly, the variable STRCDE is created to indicate whether the census housing unit is a single unit or multi-unit. If there is no unit designation for the address the variable is set to single unit.

All of the information will come from the five PerMaRCS files: the Account File, the A.C.E. Address File, the Census Address File, the A.C.E. Person File and the Census Person File. See

DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-CO-03,
Requirements for Person Matching Review and Coding System File Transfers.

Appendix A contains documentation of the new variables TESPERS, WHHNNGP, WPFIN and STRCDE. Appendix B contains layouts of the variables used in the coding, such as HICODE, etc.

TES Person Status (TESPER)

To determine a cluster's TES status check the Account File:

For each person on the A.C.E. person file:

 In clusters where TESCLUS = U, W, N, S, R or L:

 If HITES = Y and HICODE = KI, KP or DP then TESPERS = 0.

 If HITES = Y and HICODE ne KI, KP or DP then TESPERS = 1.

 If HITES = N then TESPERS = 0.

 In clusters where TESCLUS = O or I set TESPERS = 0 for all people.

For each person on the Census person file:

 In clusters where TESCLUS = U, W, N or S:

 If HITES = Y and HICODE = KE then TESPERS = 0.

 If HITES = Y and HICODE = DE then check ESAMP of the primary, which is indicated by DCENID, DCSEQ and DCPER:

 If ESAMP = 3 then TESPERS = 1.

 Else TESPERS = 0.

 If HITES = Y and HICODE ne KE or DE then TESPERS = 1.

 If HITES = N then TESPERS = 0.

 If HICODE = 'GE' and HUFIN = ' ' then TESPERS = 1.

 In clusters where TESCLUS = R or L:

 If BFUMAT = GE, GS, GC or GU then set TESPERS to 1.

 Else if BFUMAT = NE and HICODE = GE then set TESPERS to 1.

 Else set TESPERS to 0.

 In clusters where TESCLUS = O or I set TESPERS = 0 for all people.

Highest Match Code (HICODE)

For each person on the Census Person File:

If TESCLUS in ('W', 'U', 'S') and HUFIN = ' ' and HICODE = 'GE' then HICODE = 'GU'.

If TESCLUS = 'R' and BFUMAT = 'NE' and HICODE = 'GE' then HICODE = 'GU'.

Whole Household NN/GP Flag (WHHNNGP)

For each A.C.E. address on the A.C.E. address file:

If all people have HICODE = NN then WHHNNGP = 1.

Else if at least one person has HICODE = GP then do:
 Recode all people at the address to GP.
 Set WHHNNGP = 2.

Else WHHNNGP = 0.

Whole Household/Partial Household Match Code (WPFIN)

For each A.C.E. address on the A.C.E. address file:

If PPOP = 0 (address had no people in person matching) then WPFIN = 0.

If one or more person with the same CLUST, MSN and WMSN have HICODE in (M, MR, MN, MU) and if one or more person with the same CLUST, MSN and WMSN have HICODE in (NP, NR, NL, NN, NU, FP, NC) then WPFIN = 1.

If one or more people with the same CLUST, MSN and WMSN have HICODE in (NP, NR, NL, NN, NU, FP, NC) and none have HICODE in (M, MU, MR, MN) then WPFIN = 2.

If one or more people with the same CLUST, MSN and WMSN have HICODE in (M, MU, MR, MN) and none have HICODE in (NP, NR, NL, NN, NU, FP, NC) then WPFIN = 3.

All other cases WPFIN = 4.

For each census address on the census address file:

If INPS = 0 (address had no people in person matching) then WPFIN = 0.

If one or more person with the same CID have HICODE in (M, MR, MN, MU) and if one or more person with the same CID have HICODE in (CE, EE, UE, GE, DE, FE, GU) then WPFIN = 1.

If one or more people with the same CID have HICODE in (CE, EE, UE, GE, DE, FE, GU) and none have HICODE in (M, MU, MR, MN) then WPFIN = 2.

If one or more people with the same CID have HICODE in (M, MU, MR, MN) and none have HICODE in (CE, EE, UE, GE, DE, FE, GU) then WPFIN = 3.

All other cases WPFIN = 4.

Basic Street Address Code (STRCDE)

For each census housing unit on the census housing unit file:

If UNIT = '' then STRCDE = 1;

Else STRCDE = 2.

For each census person on the census person file assign the value of STRCDE of the housing unit with the same CID.

cc. DSSD Census 2000 Operations and Procedures Memorandum Series Distribution List
A.C.E. Implementation Team
Statistical Design Team Leaders

APPENDIX A - Definitions of the New Variables

TESPER (TES Person Status)

0	=	Not a TES person.
1	=	TES person.

WHHNGP (Flag for Whole Household NN or GP)

0	=	Not Whole Household NN or GP
1	=	Whole Household NN (Group Quarters/other residence)
2	=	Whole Household GP (Geocoding Error)

WPFIN (Whole Household/Partial Household Code)

0	=	House unit did not go to person matching
1	=	Partial household match
2	=	Whole household nonmatch
3	=	Whole household match
4	=	All other

STRCDE (Basic Street Address Code)

1	=	Single unit
2	=	Multi-unit

APPENDIX B - Definitions of Variables Used in the Coding

PPOP **Number of P-sample Persons**

INPS **Number of selected persons (to be on the hundred percent census unedited file (HCUF))**

CLUST **Cluster Number**

MSN **A.C.E. Map Spot Number**

WMSN **A.C.E. Within Map Spot Number**

CID **Census Unit Identification**

HITES **Current TES person flag**

Y = TES person

N = Not TES person

TESCLUS **TES cluster**

N = Not a TES cluster

U = TES cluster, Unweighted certainty

W = TES cluster, Weighted certainty

S = TES cluster, Sample selected

R = Relist

O = Out of scope

L = Hand selected L/E TES cluster

I = Ineligible for TES / no TES eligible HUs

BFUMAT Final BFU match code

Only the codes referred to in the specification are shown:

GE	=	The E-sample person is erroneously enumerated in this block cluster, because the housing unit is a geocoding error.
GS	=	The E-sample person is enumerated in a housing unit that exists in a surrounding block.
GC	=	The E-sample person is enumerated in a housing unit that exists in the sample cluster.
GU	=	The geographic work for the targeted extended search is unresolved. The field work was not done or the block number on the form was not in the surrounding blocks, in the block cluster, or on the map. It is not clear where the housing unit is located.

HICODE Current (Highest) Match Code

List of after followup (AFU) person match codes referred to in the specification:

P	=	There is not enough information collected to determine if the possible match is a match or not. The match status of the P-sample and E-sample people is unresolved.
CE	=	The E-sample person is identified as correctly enumerated from the A. C. E. person follow-up interview.
EE	=	The E-sample nonmatch is identified as erroneously enumerated from the A. C. E. person follow-up interview.
FE	=	The E-sample nonmatch is determined to be fictitious in this block cluster during the follow-up interview. The person may have existed, but could not find anyone in this cluster who knew them after talking to at least three people. The E-sample person is erroneously enumerated in the census in this block cluster.
FP	=	The P-sample person is fictitious in this block cluster. The person was interviewed in error during the person interview. This person is not included in the final P-sample.
UE	=	Not enough information is collected during the A. C. E. person follow-up interview to identify the census person as correctly or erroneously enumerated in the census. The enumeration status for the E-sample person is unresolved. This code is also used when the E-sample person is followed up to collect geographic information and that information is not collected.

NR	=	The P-sample person is identified as a resident in the block cluster on census day during the A. C. E. person follow-up interview. The P-sample person is missed in the E-sample.
NL	=	The P-sample person did not live at the sample address or in the block cluster on census day and was listed as a nonmover or outmover in error. This person is removed from the list of P-sample people, since he or she was collected during the person interview in error.
NN	=	The P-sample person is identified as a nonresident in the block cluster on census day during the A. C. E. person follow-up interview, because the person lived in group quarters on census day or had another residence where the person should have been counted on census day according to census residence rules. This person is removed from the list of P-sample people, since he or she was collected during the person interview in error.
NU	=	Not enough information is collected during the A. C. E. person follow-up interview to identify the P-sample person as a resident or nonresident in the block cluster. The residence status for the P-sample person is unresolved. This code is also used when the P-sample person is followed up to collect geographic information and that information is not collected.
KP	=	Match not attempted for the P-sample person, because (1) the name is blank or incomplete, such as "Mr. Jones" or (2) the name is not a valid name, such as "White Female" or "Donald Duck".
KI	=	Match not attempted for the P-sample person because the person has insufficient information for matching and follow-up. The name is blank or incomplete or the name is complete but the person has only one characteristic. This is a computer assigned code and these people are suppressed from view by the matchers.
KE	=	Match not attempted for the census person. The name is incomplete or not a valid name, such as "Child Jones", or "Mickey Mouse".
DP	=	The P-sample person is a duplicate of another P-sample person.
DE	=	The E-sample person is a duplicate of another E-sample person. The non E-sample person in the block cluster (i.e., E-sample indicator 2) is a duplicate of the E-sample person. The E-sample person is a duplicate of a census person in a surrounding block (i.e., E-sample indicator 3).
GP	=	The P-sample person is removed because the person interview was conducted at a housing unit that exists outside the sample cluster. The person follow-up identified this housing unit as a P-sample geocoding error.
GE	=	The E-sample person is erroneously enumerated in this block cluster, because the census housing unit is a geocoding error (i.e., counted in the block cluster in error). The E-sample person should have been enumerated elsewhere in the census. This code is only assigned to added census housing units that are determined during follow-up to be geocoded incorrectly.
MR	=	The A. C. E. person follow-up interview determined that the matched person with unresolved residence status is a resident. The person is in the P-sample

and is matched to a census person. If the census person is in the E-sample, the E-sample person is a correct enumeration.

MN = The A. C. E. person follow-up interview determined that the matched person with unresolved residence status is not a resident in this housing unit or in this block cluster. The person is no longer a P-sample person. If the match was to an E-sample person, the E-sample person is an erroneous enumeration.

MU = The A. C. E. person follow-up interview obtained no useful information to resolve the unresolved residence status for the matched person. The P-sample person's residence status is unresolved. If the match was to an E-sample person, the E-sample person's enumeration status is unresolved.